Audio Coding for Beamforming with Distributed Microphones

Matthias Pawig, Peter Vary

Institute of Communication Systems and Data Processing, RWTH Aachen University, 52056 Aachen
Email: {pawig,vary}@ind.rwth-aachen.de
Web: www.ind.rwth-aachen.de

Abstract

In this contribution we investigate the influence of audio coding on beamforming algorithms. Usually, a beamforming system would be implemented in the same device as the microphones. However, for some devices such as wireless distributed microphone arrays, or digital hearing aids, it may be necessary to transmit the microphone signals prior to combining. Since any digital signal transmission will introduce quantization noise, the subsequent beamformer processing will be influenced. In this paper, we investigate the behavior of a differential beamformer and a generalized sidelobe canceler after encoding and decoding of microphone signals. Furthermore, an example of a suitable audio coding scheme is given.

1 Introduction

In order to increase the quality of audio communications, devices like mobile telephones or digital hearing aid as well as teleconferencing systems require suppression of interfering sound sources from the target speaker signal. Therefore, many devices rely on the combination of more than a single microphone signal in conjunction with signal processing algorithms, i.e., beamforming techniques.

In many cases, it may be beneficial or necessary to separate the physical microphone from the signal processing unit, e.g., when using wireless microphones freely distributed in a conference room, or when transmitting microphone signals from one hearing aid to another. In these scenarios, the transmitted signal will be subject to encoding and decoding, introducing quantization noise which disturbs the subsequent beamforming or spatial filtering.

In this paper, we consider a system setup as seen in Figure 1. Here, two microphone signals are encoded separately or jointly at Terminal 1 and transmitted to Terminal 2, where the array processing takes place. At Terminal 2, it is possible to take two additional, local microphones into account. All microphone signals are fed into a beamforming algorithm to produce an output signal with increased audio quality.

The paper is structured as follows. In Section 2 we will describe a differential beamformer (DB) and the influence of employing two transmitted, i.e., encoded and decoded microphone signals. Section 3 will consider the behavior of a generalized sidelobe canceler (GSC) which takes into account the two transmitted microphone signals and the two local microphone signals of the remote terminal. In Section 4, an audio coding scheme will be introduced which can be used to transmit two microphone signals jointly at suitable quality, including a fixed noise shaping scheme to improve the performance of a differential beamforming stage as in Section 2. The scheme is also suitable for the subsequent employment of a generalized sidelobe canceller (GSC) structure. Using inter-signal prediction, this scheme reduces the bitrate compared to separate encoding of the two microphone signals.

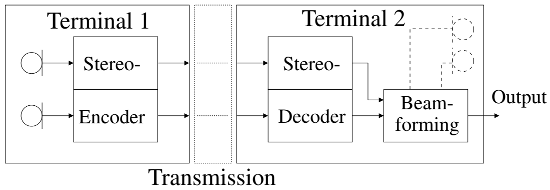

Figure 1: Block Diagram for Differential Processing After Transmission

2 Differential Beamforming of Encoded Signals

The first beamforming scheme investiated in this paper is a differential beamformer (DB) based on [1] as displayed in Figure 2. The output signals $c_B(t)$ and $c_F(t)$ of two first order differential beamformers with cardioid beam pattern are subtracted from each other.

The direction of maximum attenuation θ_{null} can be steered by adjusting the weighting factor β for the backward facing cardioid. The overall transfer function of the array after subtraction with respect to a plane-wave signal $s(t)$ with spectrum $S(\omega)$ can be given as

$$|H(\omega,\theta)| = 2|\sin\frac{\omega\frac{d}{c}(1+\cos\theta)}{2} - \beta\sin\frac{\omega\frac{d}{c}(1-\cos\theta)}{2}|, \tag{1}$$

where c is the speed of sound. Assuming the maximum attenuation is placed at angle θ_{null} in the back half plane, i.e., $90^\circ \leq \theta_{\text{null}} \leq 180^\circ$, the direction can be chosen according to

$$\theta_{\text{null}} = \arccos\frac{\beta-1}{\beta+1}, \tag{2}$$

assuming small spacing, i.e., $\omega\frac{d}{c} \ll \pi$. In time varying environments, an adaptive algorithm can be employed in order to obtain the optimum parameter β, as demonstrated in [1].

Looking at the frequency response of the differential array in (1), assuming once more a small spacing ($\omega\frac{d}{c} \ll \pi$), the frequency response can be approximated by

$$|H(\omega,\theta)| \approx \omega\frac{d}{c}|(1+\cos\theta) - \beta(1-\cos\theta)|. \tag{3}$$

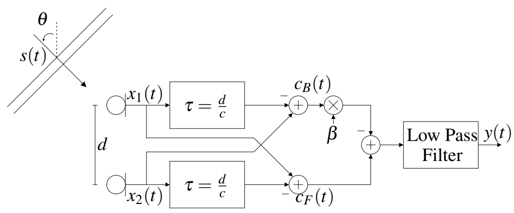

Figure 2: Diagram of steerable first-order array using a combination of forward and backward facing cardioids

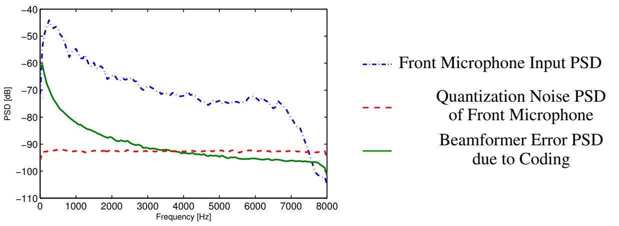

Figure 3: Effective Quantization Noise at Differential Beamformer Output

Here, a high pass characteristic, i.e., a linear increase of the magnitude with frequency can be observed. This frequency dependency has to be equalized in practice by a first-order low-pass filter as displayed in the block diagram in Figure 2.

If we now consider using this differential array after both microphone signals $x_1(t)$ and $x_2(t)$ have been transmitted, i.e., encoded and decoded, they both contain quantization noise. Since the quantization noise occurs in both microphone signals separately, it can be assumed that it is uncorrelated between both microphones and thus should be attenuated by the differential array in all but the target direction. However, due to the low-pass equalizing filter at the output of the differential beamformer, low frequency components of the quantization noise present a major problem.

Simulation results in Figure 3 demonstrate this effect. Here, a differential array has been used to suppress an interfering source in the 180° direction, i.e., $\beta = 0$. Both microphone signals have been encoded and decoded with the lattice based ADPCM codec [2] described in Section 4.1. In order to demonstrate the effect of the equalizing low-pass, the input power spectral density (PSD) is shown as well as the PSD of the quantization noise caused by audio coding of the front microphone signal. It can be observed that the audio coding effectively causes quantization noise with a white noise characteristic. Calculating the beamformer error PSD between the differential array with uncoded microphone inputs and the differential array with the encoded and decoded microphone signals, it can be seen that the effective error at the output now shows the influence of the low-pass stage, i.e., the quantization noise is amplified at lower frequencies. The exact frequency characteristic of the quantization noise will be dependent on the steering factor β. This leads to very audible distortions of the output signal. It can therefore be concluded that any audio coding for a differential array has to take this equalizing low-pass into account and compensate its characteristic. In Section 4.1, a fixed noise shaping scheme will be introduced which can be used to shape the quantization noise accordingly.

3 Influence of Audio Coding in a GSC Structure

The second beamformer system setup is a four-channel implementation of a generalized sidelobe canceler (GSC) [3],[4]. This system is fed with two microphones $x_1(k)$ and $x_2(k)$, which are located at the decoder site, i.e., do not have to be transmitted using an audio codec. Additionally, two transmitted and audio coded microphone signals $x_3(k)$ and $x_4(k)$ are fed into the GSC structure, which are subject to the audio coding quantization noises $q_3(k)$ and

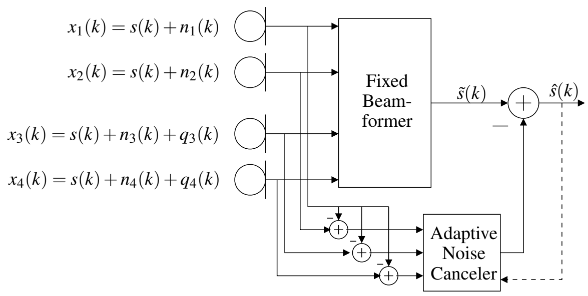

Figure 4: Noise Reduction System Signal Flow for GSC Structure with 2 Transmitted and 2 Local Microphones

$q_4(k)$. In Figure 4, a simplified block diagram for this system is shown, using the assumption that the microphone signals $x_i(k), i \in \{1,2,3,4\}$ are perfectly aligned in the target direction, i.e., the target signal component $s(k)$ occurs in all of them. The GSC system consists of a fixed beamformer stage, which in this case simply averages all of the input microphone signals, and an adaptive noise canceler, which uses the noise references constructed by subtracting the microphone signals $x_2(k)$, $x_3(k)$ and $x_4(k)$ from the first microphone signal $x_1(k)$.

When analyzing the influence of audio coding on this system, it can be observed that fixed beamformer averages the microphone input signals. Given the above assumption of identical target signal components $s(k)$ in all microphone signals, we can give the target signal power at the output as

$$S_{\text{target}} = \mathrm{E}\{(\frac{1}{4} \cdot 4 \cdot s(k))^2\} = \mathrm{E}\{s^2(k)\}, \tag{4}$$

where $\mathrm{E}\{\dots\}$ refers to the mathematical expectation. Under the further assumption of quantization noise components of equal power, the overall quantization noise power can be given as

$$Q_{\text{coding}} = \mathrm{E}\{(\frac{1}{4}q_3(k) + \frac{1}{4}q_4(k))^2\} \tag{5}$$

$$= 2\mathrm{E}\{\left(\frac{1}{4}q_3(k)^2\right)\} = \frac{1}{8}\mathrm{E}\{q_3^2(k)\}, \tag{6}$$

since the quantization noise signals are uncorrelated. Comparing the signal-to-quantization-noise-ratio of the audio coding system $\mathrm{SQR}_{\text{cod}}$ to the ratio $\mathrm{SQR}_{\text{out}}$ at the output of the fixed beamformer reveals

$$10\log_{10}(\mathrm{SQR}_{\text{out}}) = 10\log_{10}(\mathrm{SQR}_{\text{cod}}) + 9.03\,\mathrm{dB}, \tag{7}$$

i.e., a reduction of the effective quantization noise level due to the fixed beamformer, since the target signals combine constructively while the quantization noise is averaged. While this analysis holds strictly true only in case of perfect alignment of the target components, this calculation shows that a mixture of signals with a quantization error and signals without a quantization error, i.e., local signals, will improve the overall SQR at the output of the system. Similar approximations can be made for the adaptive noise canceler whose implementation goes beyond the scope of this paper but confirm that no constructive combination of quantization noise occurs.

In conclusion, in the case of a mixture of transmitted and local microphone signals, the quantization noise caused by the audio codec is effectively reduced by the beamforming system itself due to its uncorrelated nature.

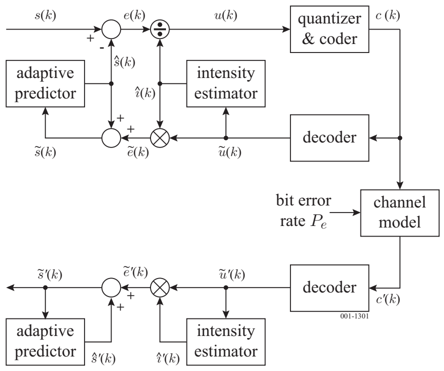

Figure 5: Block Diagram of Lattice ADPCM Codec

4 Optimized Audio Coding for Beamforming Algorithms

In order to find a suitable audio codec for subsequent beamforming, it has to be considered that the differential approaches in beamforming algorithms require a good representation of the phase of the transmitted signals. Therefore, waveform codecs are prefered to parametric encoders. Moreover, a spectrally flat characteristic is desirable to reduce the influence of the audio codec on the adaptations within the beamformer algorithms. Additionally, classic noise shaping techniques which cause coloring of the quantization noise to resemble the spectral shape of the speech signal are not as suitable, since the beamforming algorithms attenuate specific interferers and change the spectral shape of the speech signal itself.

4.1 Lattice Based Full Band ADPCM Coding

A suitable codec for encoding and decoding of single microphone signals is the full band ADPCM codec as described in [2], which uses a lattice filter for prediction.

The block diagram of the proposed codec is shown in Figure 5. The input signal $s(k)$ is predicted in a closed loop from the quantized error signal $\tilde{e}(k)$ with the help of a lattice predictor implementation. After subtraction of the prediction $\hat{s}(k)$ the residual $e(k)$ is normalized by dividing by the estimated intensity $\hat{\imath}(k)$, resulting in a normalized prediction residual $u(k)$. The normalized residual is quantized with a Lloyd-Max optimized symmetric quantizer with adjustable word length. Intensity estimation and adaptive prediction is based on the quantized signals. This means the same estimation can be performed at the receiver and no additional parameters have to be transmitted. At the sampling frequency of $f_s = 16\,\text{kHz}$, this corresponds to bit rates of $32 - 112$ kbit for one microphone signal. Since the codec is based on backward estimation, the effective quantization noise at the decoder output has white noise characteristic.

The basic implementation of this codec causes no additional signal delay. However, as proposed in the original paper [2], an FIR allpass filter $a(k)$ of length N is used to smoothen the input signal by reducing impulsive and step-wise variations and to improve the overall quality by improving the intensity estimation. This means the input signal $s(k)$ is filtered before feeding it into the codec displayed in Figure 5. At the output, the signal $\hat{s}(k)$ is filtered with the mirrored allpass filter.

In Section 2, it was observed that the low-pass equalizing filter of the differential array causes an amplification of

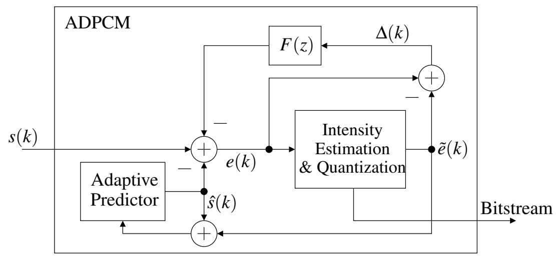

Figure 6: Encoder Side Fixed Noise Shaping

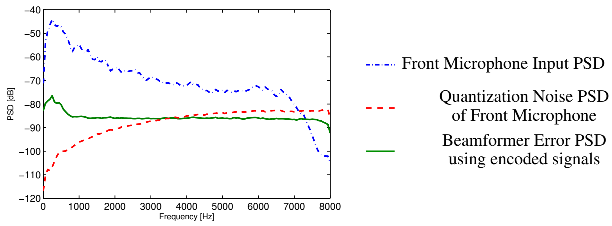

Figure 7: Effective Quantization Noise at Differential Beamformer Output Using Fixed Noise Shaping

the effective quantization error at low frequencies. The amplification can be compensated for this codec with only a change of the encoder by reconstructing the actual quantization noise as demonstrated in Figure 6. The quantization error $\Delta(k)$ is calculated by subtracting the quantized prediction residual $\tilde{e}(k)$ from the unquantized residual $e(k)$. The filter $F(z)$ can be used to shape the effective quantization noise at the output of the encoder. The effect of fixed noise shaping can be observed in Figure 7. Here, the extreme low-pass characteristic of the beamformer error PSD when using encoded microphone signals is avoided, reducing the audible distortions at the expense of slightly higher quantization noise at high frequencies. The effective high-pass characteristic of the resulting coding scheme is illustrated by the quantization noise PSD of the beamformer output. A perfectly flat characteristic of the effective quantization noise is possible only of the steering factor β is known at the encoder, e.g., in a fixed system. Otherwise, a compromise for the design of $F(z)$ has to be found.

4.2 NLMS Based Backward Inter-Signal Prediction

When transmitting two microphone signals which are spatially close, the required bitrate to transmit a second microphone signal can be reduced by employing inter-signal prediction. In this section, an additional inter-signal prediction based on normalized least mean square (NLMS) adaptation of the filter coefficients in a feedback loop is proposed.

The block diagram of the joint dual-channel encoder is shown in Figure 8. The first microphone signal $x_1(k)$ is filtered with the allpass filter $a(k)$ as explained in Section 4.1 and subsequently encoded with an implementation of the lattice ADPCM encoder which also delivers the decoded output signal $\hat{x}_1(k)$. The encoded bitstream $c_1(k)$ is transmitted to the decoder.

In order to improve the inter-predictive system, the second microphone signal is delayed by a few samples and also filtered with the allpass filter. Afterwards, the inter-prediction residual is calculated by subtracting a filtered version of the encoded and decoded microphone signal $\hat{x}_1$

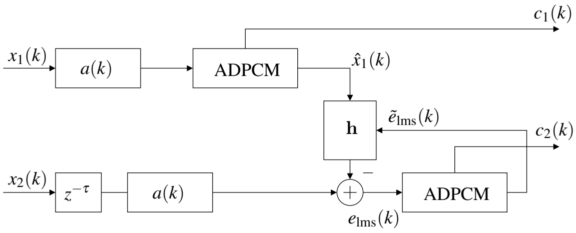

Figure 8: Encoder of the NLMS-Based Inter-Predictive Coding Approach

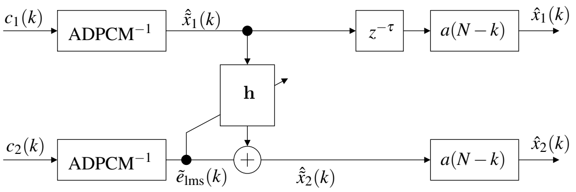

Figure 9: Decoder of the NLMS-Based Inter-Predictive Coding Approach

according to

$$e_{\mathrm{lms}}(k) = \tilde{x}_2(k) - \sum_{i=0}^{K-1} h(i,k)\hat{x}_1(k-i) \tag{8}$$

with filter length K. This inter-prediction residual is subsequently encoded with the same implementation of lattice ADPCM, where the locally decoded output signal is denoted as $\tilde{e}_{\mathrm{lms}}(k)$ and the output bitstream $c_2(k)$ is transmitted to the receiver. This decoded output is now used to calculate the filter coefficients for the next time step. Minimizing the mean square error of $\tilde{e}_{\mathrm{lms}}(k)$ leads to the update equation for the i-th prediction filter coefficient at time instant k with a stepsize $0 \leq \alpha \leq 1$:

$$h(i,k+1) = h(i,k) + \frac{\alpha}{\sum_{i=0}^{K-1} \hat{x}_1^2(k-i)} \tilde{e}_{\mathrm{lms}}(k) \cdot \hat{x}_1(k-i). \tag{9}$$

At the start of transmission, the filter coefficients $h(k,i)$ are set to zero. Since this adaptation process uses solely the output signals of the encoders, it can be performed in the same way at the decoder and the prediction coefficients do not have to be transmitted. The decoder block diagram is illustrated in Figure 9. The first step is decoding of the bitstream $c_1(k)$ to recover the decoded output signal $\hat{\tilde{x}}_1(k)$, which subsequently has to be delayed to compensate for the encoder side delay of $x_2(k)$ and filtered by the reversed allpass filter $a(N-k)$, resulting in the output signal $\hat{x}_1(k)$.

The bitstream $c_2(k)$ is used to decode the inter-prediction residual $\tilde{e}_{\mathrm{lms}}(k)$ at the receiver. Using the same adaptation process (9) as at the encoder, the filter coefficients $h(i,k)$ can be recovered. Using this filter, the output signal

$$\hat{\tilde{x}}_2(k) = \tilde{e}_{\mathrm{lms}}(k) + \sum_{i=0}^{K-1} h(i,k)\hat{\tilde{x}}_1(k-i) \tag{10}$$

is decoded and finally filtered by the reverse allpass filter $a(N-k)$ to receive the output signal $\hat{x}_2(k)$. Simulation results for the inter-signal prediction can be observed in Figure 10. The scenario is one target speaker directly in front of the array and diffuse background noise. The quantization caused by transmission with the lattice ADPCM codec

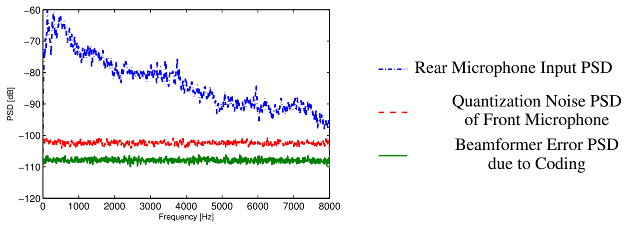

Figure 10: Input and Quantization Noise PSDs of Rear Microphone Signal

is compared to the quantization noise caused by transmission with inter-signal prediction. Comparing the noise levels, it can be seen that the inter-signal prediction achieves a signal-to-noise-ratio gain of 9 dB. This gain was consistent for a large variety of scenarios with differing number of interferers. This means that to achieve the same SNR with inter-signal prediction, the quantization word length of the rear microphone signal can be chosen to at least 1-2 bit less than using an independent transmission with lattice ADPCM.

5 Conclusions

In this paper, the influence of audio coded microphone signals at the input of two beamforming schemes was analyzed. In a differential beamformer system using only encoded input signals, the main problem is the required compensation of the equalizing low-pass filter. Analyzing the behavior of a GSC structure with both coded and uncoded microphones revealed that in this case, the beamformer inherently reduces the effective quantization noise.

Furthermore, a coding concept suitable for the transmission of two closely spaced microphones was introduced, which respects the requirements for subsequent beamforming systems. Using fixed noise shaping, the effective quantization noise can be tuned to compensate for the equalizing filter of a differential beamforming system. Using inter-signal prediction, the required bitrate for transmission to achieve reasonable audio quality at the output of a beamforming stage can be reduced.

6 Acknowledgement

Investigations on the effects on a four channel beamforming system were based on fruitful discussions, algorithms, and source code by Stefan Meier and Walter Kellermann, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany.

References

[1] G. Elko and A.-T. N. Pong, "A Simple Adaptive First-Order Differential Microphone," in *Applications of Signal Processing to Audio and Acoustics, 1995., IEEE ASSP Workshop on*, (New Paltz, NY, USA), pp. 169–172, 1995.

[2] M. Dietrich, "Performance and Implementation of a Robust ADPCM Algorithm for Wideband Speech coding with 64 kbit/s," in *Proc. International Zurich Seminar on Digital Communications*, (Zuerich, Switzerland).

[3] L. J. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *Antennas and Propagation, IEEE Transactions on*, vol. 30, pp. 27–34, Jan 1982.

[4] P. Vary and R. Martin, *Digital Speech Transmission*. John Wiley and Sons, LTD, 2006.